

# Resonant plasmonic terahertz detection in graphene split-gate field-effect transistors with lateral p-n junctions

V Ryzhii<sup>1,2,3</sup>, M Ryzhii<sup>4</sup>, M S Shur<sup>5</sup>, V Mitin<sup>1,6</sup>, A Satou<sup>1</sup>, and T Otsuji<sup>1</sup>

<sup>1</sup>*Research Institute of Electrical Communication,  
Tohoku University, Sendai 980-8577, Japan*

<sup>2</sup>*Institute of Ultra High Frequency Semiconductor Electronics,  
RAS, Moscow 117105, Russia*

<sup>3</sup>*Center for Photonics and Infrared Engineering,  
Bauman Moscow State Technical University,  
Moscow 111005, Russia*

<sup>4</sup>*Department of Computer Science and Engineering,  
University of Aizu, Aizu-Wakamatsu 965-8580, Japan*

<sup>5</sup>*Departments of Electrical, Computer,  
and Systems Engineering and Physics,  
Applied Physics, and Astronomy,  
Rensselaer Polytechnic Institute, Troy, NY 12180, USA*

<sup>6</sup>*Department of Electrical Engineering,  
University at Buffalo SUNY,  
Buffalo, NY 1460-1920, USA*

We evaluate the proposed resonant terahertz (THz) detectors on the base of field-effect transistors (FETs) with split gates, electrically induced lateral p-n junctions, uniform graphene layer (GL) or perforated (in the p-n junction depletion region) graphene layer (PGL) channel. The perforated depletion region forms an array of the nanoconstions or nanoribbons creating the barriers for the holes and electrons. The operation of the GL-FET- and PGL-FET detectors is associated with the rectification of the ac current across the lateral p-n junction enhanced by the excitation of bound plasmonic oscillations in the p- and n-sections of the channel. Using the developed device model, we find the GL-FET and PGL-FET-detectors characteristics. These detectors can exhibit very high voltage responsivity at the THz radiation frequencies close to the frequencies of the plasmonic resonances. These frequencies can be effectively voltage tuned. We show that in PL-FET-detectors the dominant mechanism of the current rectification is due to the tunneling nonlinearity, whereas in PGL-FET-detector the current rectification is primarily associated with the thermionic processes. Due to much lower p-n junction conductance in the PGL-FET-detectors, their resonant response can be substantially more pronounced than in the GL-FET-detectors corresponding to fairly high detector responsivity.

## I. INTRODUCTION

The lateral p-n-junctions in field-effect transistor (FET) structures with a graphene layer (GL) channel and split gates [1–6] can be used in new electron and optoelectronic devices. For example, the tunneling electron-transit diodes for generating terahertz radiation (THz) [7, 8], interband THz photodetectors [9–18], and injection THz lasers [19–21] have been proposed and analyzed. Recently [22], the THz emission from split-gate GL-FET structures with forward biased p-n- junctions was observed. Similarly, graphene nanoribbon (GNR) and graphene bilayer (GBL) FET structures with p-n junctions can be used in different devices [23–28]. The nonlinearity of the current-voltage characteristics of the GL [1, 4, 26–29] and GBL p-n-junctions [30–33] can be used for detecting THz radiation using the effect of the current rectification and for the THz generation using the tunneling negative differential conductivity. Since the gated regions of p- and n-types form the plasmonic cavities [34], the excitation of the plasmonic oscillations (the oscillations of the electron or hole density and the

self-consistent electric field) by incoming THz radiation can substantially affect the detector characteristics due to the plasmonic resonances. An idea to use the plasmonic properties of the electron (or hole) system in the channel of a FET for the resonant THz detection was brought forward a long time ago [34]. The key points of this idea are: (1) the resonant excitation of the plasmonic oscillations and (2) the rectification of the ac current is due to the hydrodynamics nonlinearity of the electron system dynamics in the channel. Later, the concept of FET-based THz detectors was successfully realized and such detectors based on III-V compound materials with the enhanced characteristics were manufactured and used in applications (see, for example, [35–39]). The concepts of using the combination of plasmonic effects and other types of the current nonlinearity (stronger than the hydrodynamic nonlinearity) were also considered.

Due to unique electron properties of GLs, GNR, and GBLs [40–43] and a remarkable progress in fabrication of different transistor heterostructures based on GLs, such FETs can be used for THz detectors (and frequency multipliers) with enhanced performance [44–47]. The ad-

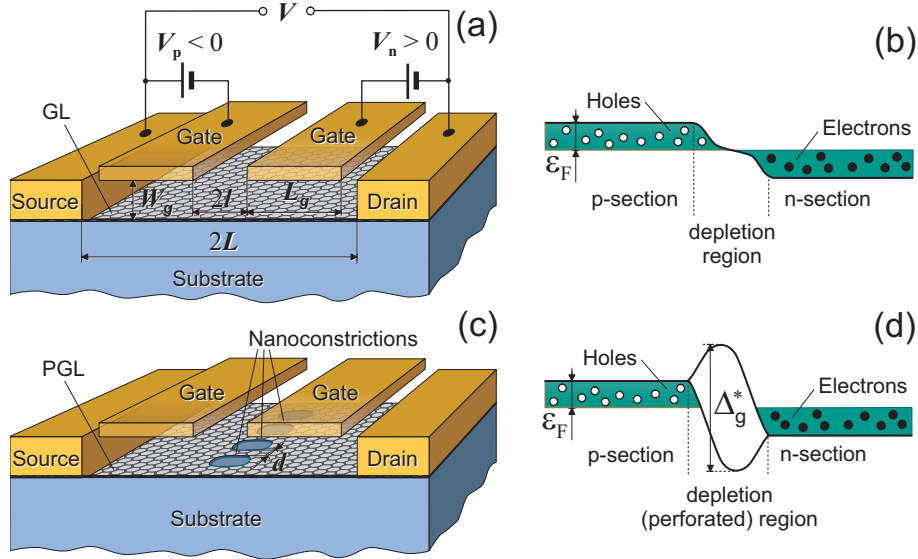


FIG. 1: Schematic views of the split-gate GL-FET (a) and PGL-FET (c) structures and their band diagrams (b) and (d), respectively (at unbiased p-n junction).

vantages of the GL-based heterostructures are associated with the enhanced mobility at room temperature (about  $110,000 \text{ cm}^2/\text{V s}$  in the epitaxially grown GLs on SIC substrate [43]). This leads to a high quality factor of the plasmonic oscillations at room temperature and high frequencies of the plasma modes in relatively long channels due to the elevated plasma-wave velocity [48]. Besides higher room-temperature quality factors compared to those for the FET-detectors based on the standard materials, the symmetry of the hole and electron energy spectra provides an opportunity to realize coherent coupled plasmonic oscillations in the p- and n-sections of the device. The vertical GL-based structures can also be used for the resonant plasmonic THz detection [49, 50]. The GNR transistor structures [23, 51] can serve the basis for the THz detectors using different operational principles [14], including the resonant absorption of THz radiation [53, 54] leading to the rectification and bolometric effects. High quality factor of the plasmonic oscillations in GL-heterostructures, opens up the prospect of realization of the room-temperature THz detectors exhibiting pronounced resonant behavior due to the plasmonic effects with very high peak responsivity and strong spectral selectivity.

In this paper, we evaluate the proposed resonant THz detector based on lateral split-gate GL-FET and split-gate FET with the GL perforated between the gates (PGL-FET) in which the channel is partitioned into the p- and n-type sections.

The operation of such detectors is enabled by the non-linearity of the p-n-junction and the excitation of coupled plasmonic oscillations in the gated channel sections. We demonstrate that the responsivity of the GL-FETs under consideration as function of the incoming THz radiation

frequency can exhibit the resonant response. However, the height of the spectral dependence responsivity peaks and their sharpness are rather moderate. This is because of relatively high conductance of the p-n junction in the GLs associated with the interband tunneling can be compared with the conductance of the channel. In contrast, in the PGL-FETs the tunneling mechanism can be effectively suppressed by appropriate choice of the energy gap in the perforated portion of the channel (perforated depletion region of the p-n junction). Apart from this, the thermionic conductance of the p-n junction is also limited by the channel nanoconstrictions. As a result, the net conductance of the p-n junction can be much smaller than that of the gated channel sections that is beneficial for the achievement of pronounced spectral selectivity and elevated peak responsivity.

## II. DEVICE MODEL

In FETs with undoped GL-channel and with the split gates, the formation of the lateral p-n junction is realized by applying voltages  $V_p$  and  $V_n$  between the source and the neighboring gate section (source gate section) and between the drain and the drain gate section. Figure 1 shows the GL- and PGL-FETs, which can be used for the THz detection and their band diagrams at properly chosen applied gate voltages.

The main feature of the PGL-FETs is the energy band gap in the depletion region caused by the GL perforation substantially affecting the conductivity, nonlinear, and plasmonic properties of the p-n-junction (as in FETs based on heterostructure made of the standard materials [55, 56]). This region can be considered as an array

of the nanoconstrictions or GNRs similar to the GNR channels of FETs [51] and bolometric detectors [52]. The source-drain voltage comprises the dc and ac signal components:  $V = V_0 + \delta V_\omega \exp(-i\omega t)$  is applied between the side contacts, where  $V_0$  is the dc bias voltage in the case of photocurrent detection and the dc signal voltage produced by the incoming radiation in the photovoltaic regime,  $\delta V_\omega$  and  $\omega$  are the amplitude and frequency of the ac signal received by an antenna.

We assume that the steady-state hole and electron densities in the GL channel,  $\Sigma_0^+$  and  $\Sigma_0^-$ , are equal to each other:  $\Sigma_0^\pm = \Sigma_0$ . If, in addition, both the gates have similar length,  $L_g$ , and due to the symmetry of the hole and electron energy spectra, the plasmonic properties of the p- and n-channel sections are similar. In this case, the plasmonic oscillations in the p- and n-channel sections can be synchronized.

If the length of the gates  $L_g$  and the net length of the channel  $L$  ( $L_g \leq L$ ) are markedly larger than the gate layers thickness  $W_g$ , one can apply the gradual channel approximation [57]. Assuming that the two-dimensional hole system (2DHS) in the p-section and two-dimensional electron system (2DES) in the n-section of the GL channel are degenerate, one can write down the following equations for the hole (in the p-section) and electron (in the n-section) density in the undoped channel  $\Sigma_0$  (at zero Dirac point):

$$\Sigma_0 = \frac{\kappa(eV_g - \varepsilon_F)}{4\pi e^2 W_g}. \quad (1)$$

Here we set  $V_p = -V_g < 0$ ,  $V_n = V_g > 0$   $e = |e|$  is the electron charge,  $\kappa$  is the dielectric constant of the gate layer material, and  $\varepsilon_F$  is the hole and electron Fermi energy in the pertinent channel sections. The Fermi energy is equal to  $\varepsilon_F = \hbar v_W \sqrt{\pi \Sigma_0}$ , where  $v_W \simeq 10^8$  cm/s is the characteristic velocity of electrons and holes in GLs, and  $\hbar$  is the Planck constant. The term with  $\varepsilon_F$  in the right-hand side of equation (1) is associated with the effect of the quantum capacitance [58]. The hole and electron systems degeneracy implies that  $\varepsilon_F \gg T$ , where  $T$  is the temperature (in the energy units).

Consequently, the small-signal variations of the hole density  $\delta \Sigma_\omega^+$  and the electron density  $\delta \Sigma_\omega^-$  caused by the ac variations of the electric potential in the channel  $\delta \varphi_\omega^\pm$  are given by

$$\delta \Sigma_\omega^\pm = \mp \frac{C}{e} \delta \varphi_\omega^\pm. \quad (2)$$

Here  $C = C_g C_{quant} / (C_g + C_{quant})$  is the net gate-GL capacitance (per unit area) accounting for the geometrical and quantum capacitances [57],  $C_g = (\kappa / 4\pi W_g)$ ,  $C_{quant} = 2e^2 \sqrt{\Sigma_0} / \sqrt{\pi \hbar} v_W$  (in GLs [58]), i.e.,  $C = (\kappa / 4\pi W_g) [1 + (\kappa \hbar^2 v_W^2 / 8e^2 W_g \varepsilon_F)]^{-1}$ .

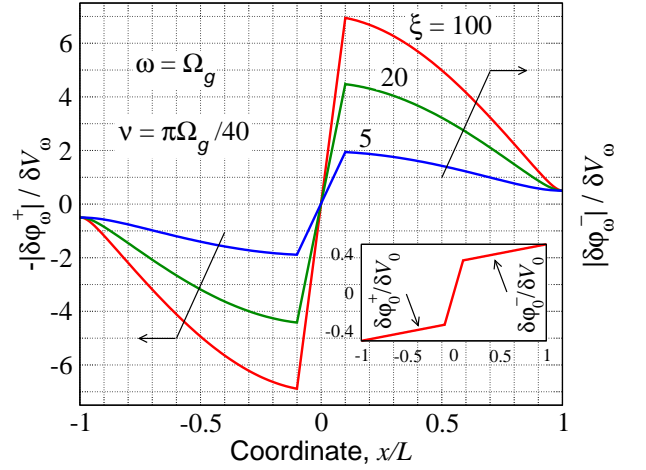


FIG. 2: Spatial distributions of the ac potential amplitudes for different  $\xi$  (schematic view) at the signal frequency  $\omega = \Omega_g$ . Inset shows an example of the spatial dependence of the potential at low frequencies ( $0 \leq \omega \ll \nu, \Omega_g$ ).

### III. THERMIONIC AND TUNNELING CONDUCTANCE AND PARAMETER OF NONLINEARITY

The ac source-drain current is determined by the admittances of the p- and n- sections of the channel and the p-n-junction. The net ac current density through the p-n-junction  $\delta j_\omega^{pn}$  (the current per unit length in the direction along the gate edges) at the signal frequency  $\omega$ , associated with the ac voltage between the source and drain contacts induced by the incoming radiation, includes the ac interband tunneling and thermionic components  $\delta j_\omega^{tunn}$  and  $\delta j_\omega^{th}$ , respectively, and the displacement component  $\delta j_\omega^{disp}$ :

$$\delta j_\omega^{pn} = \delta j_\omega^{th} + \delta j_\omega^{tunn} + \delta j_\omega^{disp}. \quad (3)$$

The latter term in Eq. (3) is given by

$$\delta j_\omega^{disp} = -i\omega c^{pn} (\delta \varphi_\omega^+|_{x=-l} - \delta \varphi_\omega^-|_{x=+l}), \quad (4)$$

where  $c^{pn} \simeq (\kappa / 2\pi^2) \ln(2L_g/l)$  is the geometrical capacitance of the lateral p-n-junction under consideration,  $2l$  being the length of the depletion region (which is comparable with the spacing between the gate edges) in the p-n-junction (see, for example, [61]),  $\delta \varphi_\omega^+|_{x=-l}$  and  $\delta \varphi_\omega^-|_{x=+l}$  are the ac electric potentials of the quasi-neutral p- and n-sections at their edges  $x = -l$  and  $x = +l$ , respectively, and the axis  $x$  is directed from source to drain.

The ac current density associated with the tunneling and thermionic processes caused by the ac signal can be presented as

$$\delta j_\omega^{th} + \delta j_\omega^{tunn} \simeq g^{pn} (\delta \varphi_\omega^+|_{x=-l} - \delta \varphi_\omega^-|_{x=+l}) = g^{pn} \delta V_\omega^{pn}. \quad (5)$$

The variation of the dc current density,  $\overline{\Delta j} = \bar{j} - j_0$ , arising from the averaging of the nonlinear components of the ac current density and the ac voltage drop across the p-n-junction over the fast oscillations (the rectified current and voltage components), is given by

$$\overline{\Delta j} = \beta^{pn} \overline{|\delta\varphi_{\omega}^+|_{x=-l} - \delta\varphi_{\omega}^-|_{x=+l}|^2} = \beta^{pn} \overline{|\delta V_{\omega}^{pn}|^2}. \quad (6)$$

In Eqs. (6) and (7),

$$g^{pn} = \left. \frac{dj_0}{d\varphi} \right|_{V=V_0}, \quad \beta^{pn} = \left. \frac{1}{2} \frac{d^2 j_0}{d\varphi^2} \right|_{V=V_0}, \quad (7)$$

where  $j_0$  is the dc current through the p-n junction at the dc source-drain voltage  $V_0$  (bias voltage or radiation induced voltage). The quantities  $g^{pn}$  and  $\beta^{pn}$  are the p-n-junction differential conductance (i.e., the real part of its net conductance  $Y_{\omega}^{pn} = g^{pn} - i\omega c^{pn}$ ) and the parameter of nonlinearity of the p-n junction current-voltage characteristics (leading to the rectification and frequency doubling effects), respectively. The latter formulas are valid when the space-charged (depletion) region of the p-n junction is sufficiently narrow, so that the characteristic time of electron and hole transit across this region is small in comparison with the inverse signal frequency  $\omega^{-1}$ . In the opposite case, for example, at sufficiently strong the reverse bias, so that  $l$  can be relatively large,  $g^{pn}$  and  $\beta^{pn}$  can be the frequency dependent (i.e.,  $g^{pn} = g_{\omega}^{pn}$  and  $\beta^{pn} = \beta_{\omega}^{pn}$ ) due to the electron and hole transit-time effects [7, 8] (see Section VIII). Considering both the interband tunneling and the thermionic contributions, one can obtain

$$g^{pn} = g^{th}(1 + \eta), \quad \beta^{pn} = \beta^{th}(1 + \zeta). \quad (8)$$

Here the first and the second terms in equation (9) describe the contributions of the tunneling and thermionic processes, respectively. Considering equations (A5), (A6), (A9), and (A10) in the Appendix A, for the GL-FETs  $g^{th} = g^{GL}$  and  $\beta^{th} = \beta^{GL}$  we arrive at

$$g^{GL} = \left( \frac{e^2}{\pi^2 \hbar} \right) \left( \frac{4T}{\hbar v_W} \right) \exp\left(-\frac{\varepsilon_F}{T}\right), \quad (9)$$

$$\beta^{GL} = \left( \frac{e^2}{\pi^2 \hbar} \right) \left( \frac{2e}{\hbar v_W} \right) \exp\left(-\frac{\varepsilon_F}{T}\right) = \frac{eg^{GL}}{2T}, \quad (10)$$

where  $T$  is the electron and hole temperature in the channel (which we equalize to the lattice temperature) in the energy units. The quantities  $\eta = \eta^{GL}$  and  $\zeta = \zeta^{GL}$  describe the relative contributions of the intersection tunneling (see the Appendix A). In the GL-FETs,

$$\eta^{GL} = \sqrt{\frac{\hbar v_W}{8\pi l^* \varepsilon_F}} \left( \frac{\varepsilon_F}{T} \right) \exp\left(\frac{\varepsilon_F}{T}\right), \quad (11)$$

$$\zeta^{GL} = \sqrt{\frac{\hbar v_W}{32\pi l^* \varepsilon_F}} \exp\left(\frac{\varepsilon_F}{T}\right) = \eta^{GL} \left( \frac{T}{2\varepsilon_F} \right), \quad (12)$$

where  $l^*$  is the effective width of the p-n junction. Deriving equations (11) and (12), we have accounted for that the GL p-n-junction built-in voltage  $V_{bi} = 2\varepsilon_F/e$ , set for simplicity the built-in electric field in the p-n-junction depletion region equal to  $E^{bi} = V_{bi}/\pi l^* = \varepsilon_F/el^*$  with  $l^* \geq l$  considered as a phenomenological parameter (see the Appendix A) and assumed that the dc voltage is created only by the rectification effect, i.e., it is small (the photo-voltaic detection).

In the PGL-FETs, the densities of the displacement, hole, and electron currents are spatially periodic in the direction along the gate edges. Considering the PGL-FETs with the period of the perforation of the depletion region  $D \ll L_g$ , we average these current densities, conductances, and the parameter of nonlinearity over short-range periodical spatial variations. Taking this into account, for the PGL-FETs the quantity  $g^{th} = g^{PGL}$  averaged over the periodic spatial variations is presented as (see the Appendix B)

$$g^{PGL} = \frac{4e^2 N b}{\pi \hbar} \exp\left(-\frac{\varepsilon_F + \Delta_g/2}{T}\right) = B \exp\left(-\frac{\Delta_g}{2T}\right) g^{GL}, \quad (13)$$

$$\beta^{PGL} \simeq \frac{e^2 J_s^{PGL}}{2T^2} \simeq \left( \frac{2e^3 N}{\pi \hbar T} \right) \exp\left(-\frac{\varepsilon_F + \Delta_g/2}{T}\right) = B \exp\left(-\frac{\Delta_g}{2T}\right) \beta^{GL}. \quad (14)$$

Here the quantity  $\varepsilon_F + \Delta_g/2$  is the activation energy of the thermionic transitions between the channel sections through the channel nanoconstrictions between the perforations (GNRs),  $\Delta_g = 2\pi \hbar v_W/d$  is the energy gap in these nanoconstrictions,  $d$  is their width,  $N$  is the number of the nanoconstrictions per unit length,  $0.5 < b < 1$  is a numerical factor, which is determined by the shape of the energy barrier in the nanoconstrictions ( $b = 1$  and  $b = 0.5$  for the sharp trapezoidal and smooth barriers, respectively), and  $B = \pi N b \hbar v_W/T$  is a small parameter describing the limitation of the current across the p-n-junction due to the nanoconstrictions. If  $N = (1 - 10) \mu\text{m}^{-1}$  and  $b = 0.75$ , at room temperature one obtains  $B \simeq 0.055 - 0.555$ .

Since the incorporation of the nanoconstrictions is aimed to substantially decrease the p-n-junction conductance in comparison with the conductances of the gated sections of the channel, we assume in the following that the length of GNRs  $l^*$  and their energy gap  $\Delta_g$  are sufficiently large, so that the hole and electron tunneling through them is suppressed. In line with this, for PGL-FETs we set  $\eta^{PGL}, \zeta^{PGL} \ll 1$  (see the Appendix B).

#### IV. LOW-FREQUENCY GL-FET AND PGL-FET DETECTOR RESPONSIVITY

The current (in A/W units) and voltage (in V/W units) detector responsivities,  $R_\omega$  and  $\mathcal{R}_\omega$ , to the incoming signals with the frequency  $\omega$  are defined as

$$R_\omega = \overline{\Delta J}/SI_\omega, \quad \mathcal{R}_\omega = \overline{\Delta V}/SI_\omega. \quad (15)$$

Here  $\overline{\Delta J} = \overline{\Delta j}D$  is the variation of the dc current associated with the current rectified component,  $D$  is the lateral size of the device in the direction along the gate edges,  $\overline{\Delta V}$  is the rectified dc voltage between the side contacts induced by the received signals, which is somewhat smaller than the rectified voltage component across the p-n junction  $\overline{\Delta V}^{pn}$ , hence, in the photovoltaic regime,  $V_0 = \overline{\Delta V}$ . The quantities  $S = \lambda_\omega^2 G/4\pi$ ,  $I_\omega$  and  $\lambda_\omega = 2\pi c/\omega$  are the wavelength of the incident THz radiation and its intensity, the intensity, where  $c$  is the speed of light in vacuum and  $G \sim 1.5$  is the antenna gain. The quantities  $(\delta V_\omega)^2$  and  $I_\omega$  are related to each other as  $(\delta V_\omega)^2 = 4\lambda_\omega^2 I_\omega/\pi c$ .

The quantities  $\overline{\Delta J}$  and  $\overline{\Delta V}$  are proportional to  $|\delta V_\omega^{pn}|^2$ , which, in turn, is determined by the spatial distributions of the ac potentials  $\delta\varphi_\omega^\pm = \delta\varphi_\omega^\pm(x)$ . The latter is found taking into account that

$$\delta\varphi_\omega^\pm|_{x=\mp L} = \mp \frac{\delta V_\omega}{2}, \quad (16)$$

At relatively low signal frequencies ( $\omega \rightarrow 0$ ), the hole and electron densities as well as the electric potential manage to follow the variations of the potential at the contacts. In this case, the ac potentials in the gated channel sections  $\delta\varphi_\omega^\pm = \varphi_\omega^\pm(x)$  at  $l \leq |x| \leq L$  are linear functions of the coordinate  $x$  along the channel:

$$\delta\varphi_0^\pm = \mp \frac{\delta V_0}{2(1+r_0)} \left[ 1 \mp r_0 \frac{(x \pm l)}{L_g} \right] \quad (17)$$

with

$$\delta\varphi_0^\pm|_{x=\mp l} = \mp \frac{\delta V_0}{2(1+r_0)}. \quad (18)$$

Here  $r_0 = 2L_g g^{pn}/\sigma_0$  is the ratio of the p- and n- junction differential conductance  $g^{pn}$  and the net dc conductance of the p- and n-sections  $\sigma_0/2L_g$  ( $L_g \simeq L - l$ , see Fig. 1) with

$$\sigma_\omega = \frac{e^2 \varepsilon_F}{\pi \hbar^2} \frac{i}{(\omega + i\nu)}, \quad (19)$$

so that  $\sigma_0 = (e^2 \varepsilon_F/\pi \hbar^2 \nu)$ , where  $\nu$  is the electron and hole collision frequency in the pertinent section associated with the scattering on impurities and acoustic phonons. A schematic view of the potential distribution

at low frequencies is shown (for  $r_0 = 0.5$ ) in the inset in Fig. 2.

As a result, the quantities  $|\delta V_0^{pn}|^2$  and  $\overline{\Delta j}$  given by Eq. (6) are equal to

$$|\delta V_\omega^{pn}|^2 = \frac{(\delta V_\omega)^2}{2(1+r_0)^2}, \quad \overline{\Delta j} = \frac{D\beta^{pn}(\delta V_\omega)^2}{2(1+r_0)^2}. \quad (20)$$

As seen from Eq. (20), it is desirable that the quantity  $r_0 \ll 1$ , i.e., the conductance of the gated sections of the channel markedly exceeds the differential conductance of the p-n-junction,

Equations (15) and (20) yield the following formula for the low-frequency current responsivity:

$$R_0^J = \frac{8\beta^{pn}}{cG(1+r_0)^2} = \frac{4e^2 J_S(1+\zeta)}{cGT^2(1+r_0)^2}. \quad (21)$$

Considering equation (21) and using equations (9), (10), and (12) for  $\beta_s^{GL}$  and  $\zeta^{GL}$ , we find for the low-frequency current responsivity of the GL-FET detectors

$$R_0^{GL} = \frac{0.0118}{G} \frac{eD}{\hbar v_W} \frac{(1+\zeta^{GL})}{(1+r_0^{GL})^2} \exp\left(-\frac{\varepsilon_F}{T}\right), \quad (22)$$

The low-frequency current responsivity of the PGL-FET detectors is given by

$$R_0^{PGL} = \frac{0.0372}{G} \frac{eDN}{T} \frac{(1+\zeta^{PGL})}{(1+r_0^{PGL})^2} \exp\left(-\frac{\varepsilon_F + \Delta_g/2}{T}\right) \simeq \frac{0.0372}{G} \frac{eDN}{T} \exp\left(-\frac{\varepsilon_F + \Delta_g/2}{T}\right). \quad (23)$$

In equations (21) and (22),  $r_0^{GL} = 2L_g g^{GL}(1+\zeta^{GL})/\sigma_0 \leq 1$  and  $r_0^{PGL} = 2L_g g^{PGL}(1+\zeta^{PGL})/\sigma_0 \ll 1$ , respectively, with  $\zeta^{GL} > 1$  (or  $\zeta^{GL} \gg 1$ ) and  $\zeta^{PGL} \ll 1$ . Due to a relatively large value of  $\zeta^{GL}$  (because of the contribution of the tunneling in the GL-FETs) and large ratio of the "thermionic" exponential factors in equations (22) and (23), the current responsivity of the GL-FET detectors substantially exceeds that of the PGL-FET detectors.

In the photo-voltaic regime, the current density  $\overline{\delta j}$  is compensated by the dc current caused by the induced dc voltage  $\overline{\Delta V}$ . Considering this and taking into account that the net channel dc resistance  $r_{ch} = 2L_g/\sigma_0 + 1/g^{pn} = (1+r_0)/g^{pn}$ , we obtain

$$\frac{g^{pn}\overline{\Delta V}}{(1+r_0)} = \frac{\beta^{pn}(\delta V_0)^2}{2(1+r_0)^2}. \quad (24)$$

Hence, the dc voltage induced by the THz radiation between the source and drain contact is equal to

$$\overline{\Delta V} = \frac{2\beta^{pn}}{g^{pn}(1+r_0)} \frac{\lambda_\omega^2 I_\omega}{\pi c}. \quad (25)$$

Considering this and using equations (20), for the voltage responsivity we obtain following formula valid for the detectors of both types:

$$\mathcal{R}_0 \simeq \frac{8\beta^{pn}}{g^{pn}cG(1+r_0)}. \quad (26)$$

Consequently,

$$\mathcal{R}_0^{GL} \simeq \overline{\mathcal{R}}_0^{GL} \left( \frac{1+\zeta^{GL}}{1+\eta^{GL}} \right) \cdot \mathcal{P}_0^{GL}, \quad (27)$$

with  $\overline{\mathcal{R}}_0^{GL} = \frac{4e(1+r_0^{GL})}{cGT}$  and

$$\mathcal{R}_0^{PGL} \simeq \overline{\mathcal{R}}_0^{PGL} \left( \frac{1+\zeta^{PGL}}{1+\eta^{PGL}} \right) \cdot \mathcal{P}_0^{PGL} \quad (28)$$

with  $\overline{\mathcal{R}}_0^{PGL} = \frac{4e(1+r_0^{PGL})}{cGT}$ , where  $\mathcal{P}_0^{GL/PGL} = (1+r_0^{GL/PGL})^{-2}$ .

As follows from equation (27), the enhancement of the tunneling probability (an increase of the quantity  $\eta^{GL}$  in the GL-FETs results in a decrease in the  $\mathcal{R}_0^{GL}$ . Thus the tunneling through the p-n junction plays a negative role. This is because the tunneling component of the p-n junction conductance can exceed the thermionic component at small  $l^*$  providing relatively low p-n junction resistance. A rise of the tunneling conductivity  $g^{tunn}$  is accompanied by a relatively slow increase in the tunneling nonlinearity parameter  $\beta^{tunn}$ .

At the realistic parameters,  $\eta^{GL} \gg 1$ , so that  $\zeta^{GL} = (T/2\varepsilon_F)\eta^{GL}$ . Taking into this condition, for the low-frequency voltage responsivities we obtain

$$\mathcal{R}_0^{GL} \simeq \frac{4e}{cGT} \left( \frac{T/2\varepsilon_F}{1+r_0^{GL}} \right), \quad \mathcal{R}_0^{PGL} \simeq \frac{4e}{cGT} \frac{1}{1+r_0^{PGL}}. \quad (29)$$

The estimates for the low-frequency responsivities of the GL-FET and PGL-FET detectors at room temperature,  $\varepsilon_F = 50$  meV,  $r_0^{GL} \leq 1$ ,  $r_0^{PGL}, \eta^{PGL}, \zeta^{PGL} \ll 1$  (in the PGL-FETs with sufficiently long nanoconstrictions and properly chosen their the energy gap) yield  $\mathcal{R}_0^{GL} \leq 1 \times 10^3$  V/W ( $r_0^{GL} \leq 1$ ) and  $\mathcal{R}_0^{PGL} \simeq 5 \times 10^3$  V/W, respectively.

## V. PLASMONIC OSCILLATIONS IN GL- AND PGL-FETS

At higher frequencies, the reactive components of the two-dimensional hole system (2DHS) and two-dimensional electron system (2DES) ac conductivity  $\sigma_\omega$  (which correspond to the inertia of the electron and hole motion [61]), the p-n-junction conductance, and the gate-channel capacitances become more important because they can substantially affect the spatio-temporal distributions of the ac potential in the different portions

of the channel. In this situation, one needs to determine  $\delta\varphi_\omega^\pm|_{\mp} = \delta\varphi_\omega^\pm|_{\mp}(x)$  accounting for these factors. We find the spatial distributions of  $\delta\varphi_t^\pm(x)$  in the GL plane (along the axis  $x$ ) and, in particular,  $\delta\varphi_\omega^\pm|_{\mp}$ , from the linearized kinetic [48] or the hydrodynamic equations for the gated 2DHS and 2DES [62, 63] (adopted for the energy spectra of the electrons and holes in GLs) coupled with the Poisson equation in the gradual channel approximation, i.e., equation (2). As a result, we arrive at the following system equations:

$$\frac{d^2\delta\varphi_\omega^\pm}{dx^2} + \frac{\omega(\omega+i\nu)}{s^2}\delta\varphi_\omega^\pm = 0, \quad (30)$$

Here  $s = \sqrt{\frac{e^2\varepsilon_F}{\pi C\hbar^2}} \geq \sqrt{\frac{4e^2W_g\varepsilon_F}{\kappa\hbar^2}} \propto \Sigma_0^{1/4}$  is the characteristic plasma-wave velocity in the gated GL structures [48, 61]. The boundary conditions for Eqs. (29) and (30) are given by Eq. (16) (at the side contacts  $x = \mp L$  and the following conditions at  $x = \mp l$ :

$$-\sigma_\omega \frac{d\delta\varphi_\omega^\pm}{dx} \Big|_{x=\mp l} = \pm \delta j_\omega^{pn}, \quad (31)$$

where  $\delta j_\omega^{pn}$  is given by equation (3) and the following ones and we  $l = L - L_g \ll L_g, L$ . Writing down equation (30), we have disregarded nonlinear terms associated with the nonlinearity of the 2DES and 2DHS dynamics (in particular, the nonlinearity of the hydrodynamic equations for the 2DHS and 2DES - we have used their linearized versions!). This is because, we focus on the GL-FET operation associated with another and stronger nonlinearity (nonlinearity of the p-n-junction current-voltage characteristics).

Equation (30) with the boundary conditions (16) and (31) yield the following formulas for the spatial distributions of ac potentials  $\delta\varphi_\omega^\pm$ :

$$\delta\varphi_\omega^\pm = \mp \frac{\delta V_\omega}{2} \left\{ \cos[\mathfrak{x}_\omega(x \pm L)] + \frac{\xi_\omega \sin(\mathfrak{x}_\omega L_g) - \cos(\mathfrak{x}_\omega L_g)}{\xi_\omega \cos(\mathfrak{x}_\omega L_g) + \sin(\mathfrak{x}_\omega L_g)} \sin[\mathfrak{x}_\omega(x \pm L)] \right\} \quad (32)$$

for the range  $-L \leq x \leq -l$  for  $\delta\varphi_\omega^+$ , and  $l \leq x \leq L$  for  $\delta\varphi_\omega^-$ . Here the wavenumber  $\mathfrak{x}_\omega$  and the characteristic plasma frequency for the gated 2DES and 2DHS  $\Omega_g$  are given by

$$\mathfrak{x}_\omega = \frac{\pi\sqrt{\omega(\omega+i\nu)}}{2\Omega_g L_g}, \quad \Omega_g = \frac{\pi s}{2L_g} = \sqrt{\frac{\pi e^2\varepsilon_F}{4L_g^2 C\hbar^2}}. \quad (33)$$

The parameter  $\xi_\omega = \sigma_\omega \mathfrak{x}_\omega / 2Y_\omega^{pn} = (\sigma_\omega \mathfrak{x}_\omega / 2g^{pn} / (1 - i\omega\tau))$ , where  $\tau = c^{pn}/g^{pn}$  is the p-n junction recharging time, characterizes the ratio of the channel and p-n junction conductivities. It is given by

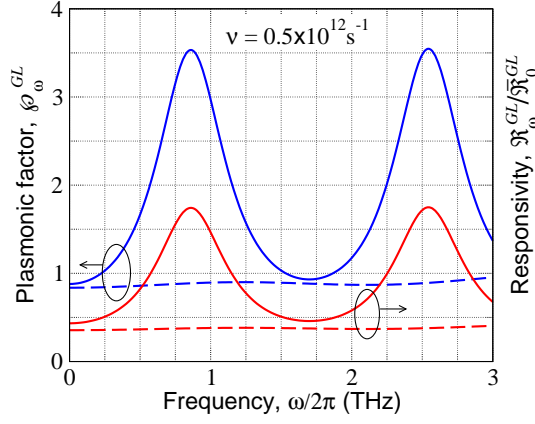


FIG. 3: Frequency dependences of plasmonic factor  $\mathcal{P}_\omega^{GL}$  and normalized GL-FET detector responsivity  $\mathcal{R}_\omega^{GL}/\mathcal{R}_0^{GL}$  for  $\nu = 0.5 \times 10^{12} \text{ s}^{-1}$  and  $L_g = 325 \text{ nm}$ : solid lines-  $l^* = 25 \text{ nm}$ ,  $W_g/\kappa = 1 \text{ nm}$ , and  $\Omega_g/2\pi = 0.86 \text{ THz}$  and dashed lines -  $l^* = 10 \text{ nm}$ ,  $W_g/\kappa = 2.5 \text{ nm}$ , and  $\Omega_g/2\pi = 1.17 \text{ THz}$ .

$$\xi_\omega = i\sqrt{\frac{\omega}{\omega + i\nu}} \frac{\xi}{(1 - i\omega\tau)}. \quad (34)$$

At low frequencies, equation (32) reduces to equation (16).

For the GBL-FETs, the quantities  $\xi = \xi^{GL}$  and  $\tau = \tau^{GL}$  are equal to

$$\xi^{GL} = \frac{\pi}{16(1 + \eta^{GL})e} \sqrt{\frac{\kappa\varepsilon_F}{W_g}} \frac{\hbar v_W}{T} \exp\left(\frac{\varepsilon_F}{T}\right), \quad (35)$$

$$\tau^{GL} = \frac{\kappa\hbar}{8(1 + \eta^{GL})T} \left(\frac{\hbar v_W}{e^2}\right) \ln\left(\frac{2L_g}{l}\right) \exp\left(\frac{\varepsilon_F}{T}\right). \quad (36)$$

The parameter  $\eta^{GL}$  given by in equations (11) varies in a wide range depending on the effective widths of the depletion region  $2l^*$ , the Fermi energy  $\varepsilon_F$ , and the temperature  $T$ . At  $T = 300 \text{ K}$ , setting  $l^* = 5 - 25 \text{ nm}$  and  $\varepsilon_F = 25 - 75 \text{ meV}$ , we obtain  $\eta \simeq 0.54 - 15.85$ . When  $\eta^{GL} \gg 1$ , from equation (34) we arrive at  $\xi^{GL} \simeq \frac{\pi^{3/2}\sqrt{\hbar v_W}}{2^{5/2}e} \sqrt{\frac{\kappa l^*}{W_g}} \simeq 0.648 \times \sqrt{\frac{\kappa l^*}{W_g}}$ . For  $\kappa = 4 - 20$  and  $l^*/W_g = 0.5 - 5.0$ , one obtains  $\xi^{GL} \simeq 0.9 - 6.5$ . As seen in the following, the latter parameter essentially determines the character of the plasmonic effect.

Due to the smallness of  $\tau$ , the factor  $(1 - i\omega\tau)^{-1}$  in Eq. (33) can be significant at very high frequencies. Indeed, at  $L_g = 250 \text{ nm}$ ,  $l^* = 10 \text{ nm}$ ,  $W_g = 10 \text{ nm}$ ,  $\kappa = 4 - 20$ , and  $\varepsilon_F = 50 \text{ meV}$ , one obtains  $\tau^{GL} \simeq (0.05 - 0.25) \times 10^{-13} \text{ s}$ . Hence, even for  $\omega/2\pi = 5 \text{ THz}$ , one obtains  $\omega\tau^{GL} \simeq 0.15 - 0.75$ .

For PGL-FETs, the quantities  $\xi = \xi^{PGL}$  and  $\tau = \tau^{PGL}$  for the parameter  $\xi_\omega$  given by equation (33) [compare with equations (35) and (36)], are as follows:

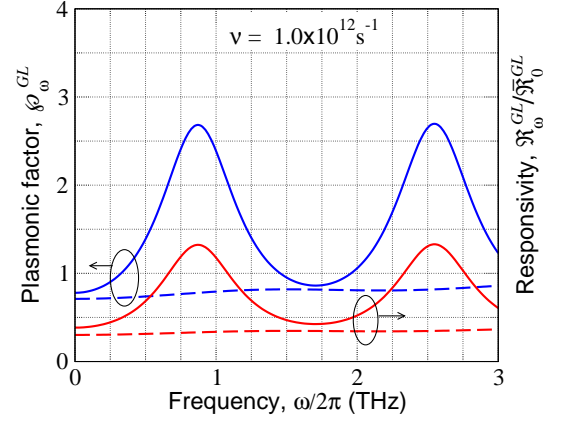


FIG. 4: The same as in figure 2, but for  $\nu = 1.0 \times 10^{12} \text{ s}^{-1}$ .

$$\xi^{PGL} = \frac{\pi}{16Be} \sqrt{\frac{\kappa\varepsilon_F}{W_g}} \frac{\hbar v_W}{T} \exp\left(\frac{\varepsilon_F + \Delta_g/2}{T}\right), \quad (37)$$

$$\tau^{PGL} = \frac{\kappa\hbar}{8BT} \left(\frac{\hbar v_W}{e^2}\right) \ln\left(\frac{2L_g}{l}\right) \exp\left(\frac{\varepsilon_F + \Delta_g/2}{T}\right). \quad (38)$$

Setting  $\kappa = 4$ ,  $W_g = 8 \text{ nm}$ ,  $d = 10 \text{ nm}$  ( $\Delta_g \simeq 200 \text{ meV}$ ),  $N = 10 \mu\text{m}^{-1}$ ,  $b = 0.75$ ,  $L_g/l = 10$ ,  $\varepsilon_F = 25 - 75 \text{ meV}$ , we obtain  $\xi^{PGL} \simeq 116.47 - 1487$  and  $\tau^{PGL} \simeq (4.35 - 32.17) \times 10^{-12} \text{ s}$ . Thus,  $\xi^{PGL}$  is fairly large. Apart from this, the recharging time,  $\tau^{PGL}$ , of the perforated p-n junction in the PGL-FETs is relatively long (in comparison with  $\tau^{GL}$ ) due to a small p-n junction conductance. Figure 2 schematically shows the amplitudes of the ac potential as a function of the coordinate  $x$  along the channel in the devices under consideration with different values of the parameter  $\xi$  ( $\xi^{GL}$  or  $\xi^{PGL}$ ). The radiation frequency  $\omega$  is chosen to be equal to  $\Omega_g$  with  $\nu = \pi\Omega_g/40$  and  $l/L = 10$ . This corresponds to the maximum span of the plasmonic oscillations. One can see from Fig. (2) that an increase in  $\xi$  results in a substantial increase in the plasmonic oscillations, which, in turn, provides an elevated ac potential drop across the the depletion region. An example of the potential distribution shown in the inset in Fig. 2 and described by equations (16) and (17) corresponds to relatively low frequencies ( $0 \leq \omega \ll \nu, \Omega$ ).

## VI. EFFECT OF PLASMONIC OSCILLATIONS ON THE GL-FET AND PGL-FET DETECTORS RESPONSIVITY

Equations (31) and (6) yield [compare with equation (20)]

$$|\overline{\delta V_\omega^{pm}}|^2 = \frac{(\delta V_\omega)^2}{2} \mathcal{P}_\omega, \quad \overline{\Delta J} = \frac{D\beta^{pm}(\delta V_\omega)^2}{2} \mathcal{P}_\omega, \quad (39)$$



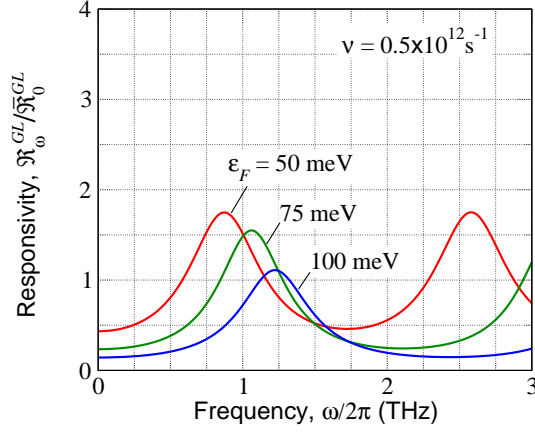


FIG. 5: Spectral dependences of normalized GL-FET detector responsivity  $\mathcal{R}_\omega^{GL}/\mathcal{R}_0^{GL}$  calculated for different hole and electron Fermi energies ( $\nu = 0.5 \times 10^{12} \text{ s}^{-1}$ ,  $L_g = 325 \text{ nm}$ ,  $l^* = 25 \text{ nm}$ ,  $W_g/\kappa = 1 \text{ nm}$ ).

where the factor

$$\mathcal{P}_\omega = \left| \frac{\xi_\omega}{\sin(\mathfrak{x}_\omega L_g) + \xi_\omega \cos(\mathfrak{x}_\omega L_g)} \right|^2 \quad (40)$$

describes the effect of plasmonic oscillations. In the range of low frequencies  $\omega$ , the plasmonic factor  $\mathcal{P}_\omega$  tends to  $\mathcal{P}_0 = (1 + r_0)^{-2}$ , and the quantities in equation (38) coincide with those given by equation (20). As a result, for the PGL-FET detectors we can use the following formula:

$$\mathcal{R}_\omega^{GL} \simeq \overline{\mathcal{R}}_0^{GL} \mathcal{P}_\omega^{GL}, \quad \mathcal{R}_\omega^{PGL} \simeq \overline{\mathcal{R}}_0^{PGL} \mathcal{P}_\omega^{PGL}. \quad (41)$$

As follows from equations (40) and (41), the plasmonic factors  $\mathcal{P}_\omega^{GL}$  and  $\mathcal{P}_\omega^{PGL}$ , as well as the pertinent responsivities  $\mathcal{R}_\omega^{GL}$  and  $\mathcal{R}_\omega^{PGL}$  are generally oscillatory functions of  $\omega/\Omega_g$ . The oscillations of  $\mathcal{P}_\omega^{GL}$  and  $\mathcal{R}_\omega^{PGL}$  can be pronounced, i.e., the pertinent maxima are high and sharp if the quality factor  $Q = 4\Omega_g/\pi\nu$  is sufficiently large.

Figures 3 and 4 show the frequency dependences of the plasmonic factor  $\mathcal{P}_\omega^{GL}$  calculated using equations (40) and (41) with equations (33) - (36) for  $\nu = 1.0 \times 10^{12} \text{ s}^{-1}$ , and  $\nu = 0.5 \times 10^{12} \text{ s}^{-1}$ . It is assumed that  $L_g = 325 \text{ nm}$ ,  $\varepsilon_F = 50 \text{ meV}$ ,  $l^* = 25$  and  $W_g/\kappa = 1.0$  for solid lines and  $l^* = 10 \text{ nm}$  and  $W_g/\kappa = 25 \text{ nm}$  for dashed lines. In particular, if  $\kappa = 10$  (SiC or MoS<sub>2</sub> gate layer) and  $\kappa = 20$  (HfO<sub>2</sub> gate layer), the solid lines with pronounced maxima, correspond to  $W_g = 10$  and  $W_g = 20 \text{ nm}$ , respectively. The solid lines in figures 2 and 3 can also be attributed to the devices with hBN or SiO<sub>2</sub> gate layers of the thickness  $W_g \simeq 4 \text{ nm}$ .

Figure 5 shows the normalized GL-FET detector responsivity as a function of the signal frequency calculated for different Fermi energies of holes and electrons, i.e., their different densities for the same other parameters corresponding to the solid lines in figure 2.

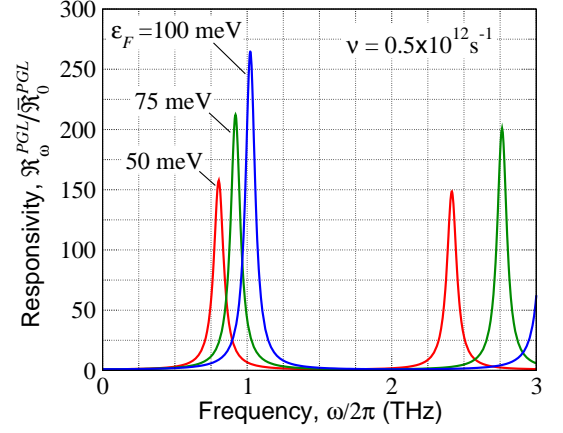


FIG. 6: Spectral dependences of normalized PGL-FET detector responsivity  $\mathcal{R}_\omega^{PGL}/\mathcal{R}_0^{PGL}$  with  $\Delta_g = 200 \text{ meV}$  and different hole and electron Fermi energies at  $\nu = 0.5 \times 10^{12} \text{ s}^{-1}$ ,  $L_g = 325 \text{ nm}$ ,  $W_g/\kappa = 1 \text{ nm}$ , and  $\kappa = 4$ .

The pronounced resonant peaks in figures 3 - 5 correspond to the frequencies  $\omega \simeq (2n - 1)\Omega_g$ , where  $n = 1, 2, 3, \dots$  is the plasmonic mode index.

Larger values of the quality factor  $Q$  correspond to higher peaks of the plasmonic factor and the GL-FET detector responsivity. In particular, solid lines correspond from  $Q = 13.76$  (in figure 3) and  $Q = 6.88$  (in figure 4). However, the comparison of the resonant peaks in figures 3 and 4 shows that an increase in the quality factor (twice) leads only to a modest increase in the peak heights. The point is that the spectral behavior of  $\mathcal{P}_\omega^{GL}$  and  $\mathcal{R}_\omega^{GL}$  is also determined by the parameter  $\xi^{GL}$  characterizing the ratio of the p-n junction and the channel resistances. In particular, the data in Figs. 3 and 4 correspond to fairly moderate values of  $\xi^{GL} \propto \sqrt{\kappa l^*/W_g}$ :  $\xi^{GL} = 2.189$  for solid lines and  $\xi^{GL} = 1.296$  for dashed lines. This is the manifestation of the effect of relatively high conductance of the p-n junction in GL-FETs, which drops when  $l^*$  decreases. The role of this effect also reveals in the fact that  $\mathcal{R}_0^{GL}/\overline{\mathcal{R}}_0^{GL} < 1$  seen in figures 2 - 4.

In the PGL-FETs the parameters  $\xi = \xi^{PGL}$  and  $1/r_0^{PGL}$  can be very large due to relatively small the p-n junction conductance  $g^{PGL}$  and negligible contribution of the tunneling mechanism ( $\eta^{PGL}, \zeta^{PGL} \ll 1$ ).

Figures 6 and 7 show the spectral dependences of the normalized responsivity  $\mathcal{R}_\omega^{PGL}/\mathcal{R}_0^{PGL}$  of PGL-FET detectors with different hole and Fermi energy  $\varepsilon_F$  the energy gap in the constrictions  $\Delta_g = 200 \text{ meV}$  ( $d = 10 \text{ nm}$ ), and their density  $N = 10 \mu\text{m}^{-1}$  calculated using equations (39) and (40) with equations (36) and (37). It is assumed that  $\nu = 0.5 \times 10^{12} \text{ s}^{-1}$  and  $\nu = 1.0 \times 10^{12} \text{ s}^{-1}$ ,  $L_g = 325 \text{ nm}$ ,  $L_g/l = 10$ ,  $W_g/\kappa = 1 \text{ nm}$ ,  $\kappa = 4$ , and  $b = 0.75$  ( $B = 0.555$ ). The lines for  $\varepsilon_F = 50 \text{ meV}$  in figure 5 and in figure 6 correspond to the same parameters of the gated sections in the PGL- and GL-FETs. Comparing figures 6 and 7, one can see that the height of the resonant responsivity peaks markedly drops with



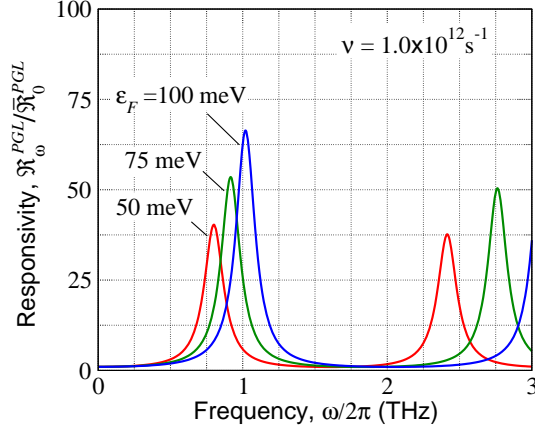


FIG. 7: The same as in figure 6 but for  $\nu = 1.0 \times 10^{12} \text{ s}^{-1}$  (different vertical axis scale).

increasing hole and electron collision frequency  $\nu$ . This is also seen from figure 8, which demonstrates how the spectral dependence of the PGL-FET-detector normalized  $\mathcal{R}_\omega^{PGL}/\mathcal{R}_0^{PGL}$  varies when the  $\nu$  changes. As seen, when  $\nu$  increases threefold, the height of the first resonant peak ( $n = 1$ ) drops nine times. Such a roll-off corresponds to  $\mathcal{R}_\omega^{PGL}/\mathcal{R}_0^{PGL} \simeq Q^2 \propto \nu^{-2}$  [see the solid and dashed lines in the inset in figure 8].

## VII. COMPARISON OF GL- AND PGL-FET DETECTORS

The variation of  $\Delta_g$  in rather wide range  $\Delta_g = 100 - 300 \text{ meV}$  at fixed  $\epsilon_F$  does not lead to a marked change in the spectral dependences of  $\mathcal{R}_\omega^{PGL}/\mathcal{R}_0^{PGL}$ .

It is clearly seen that the plasmonic response of the PGL-FETs is much more pronounced than that of the GL-FETs (at similar parameters of the gated channel sections).

Equation (40) for the plasmonic factor at the frequencies  $\omega = (2n - 1)\Omega_g$  when  $Q > 1$  yields

$$\mathcal{P}_{(2n-1)\Omega_g} \simeq \frac{\xi^2 Q^2}{(\xi + Q)^2 + (2n - 1)^2 Q^4 (\pi \nu \tau / 4)^2}. \quad (42)$$

At  $\omega = 2n\Omega_g$ , one obtains

$$\begin{aligned} \mathcal{P}_{2n\Omega_g} &\simeq \frac{\xi^2 Q^2}{Q^2 [\xi + 2n(\pi \nu \tau / 4)]^2 + 1} \\ &\simeq \frac{\xi^2}{[\xi + 2n(\pi \nu \tau / 4)]^2}. \end{aligned} \quad (43)$$

Hence, for both GL- and PGL-FETs,  $\mathcal{P}_{2n\Omega_g} < 1$ . Taking into account equation (40), for the responsivity at  $\omega = (2n - 1)\Omega_g$ , one can obtain

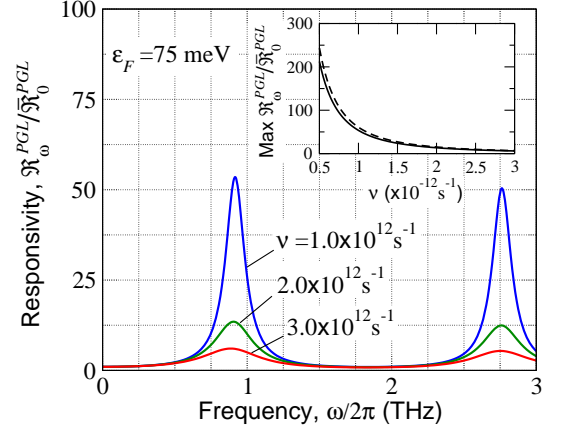


FIG. 8: Transformation of spectral dependence of PGL-FET-detector normalized responsivity  $\mathcal{R}_\omega^{PGL}/\mathcal{R}_0^{PGL}$  with varying collision frequency  $\nu$  and  $\epsilon_F = 75 \text{ meV}$ , other parameters are the same as in figures 6 and 7. Solid line in inset gives the maximum (peak) value of  $\mathcal{R}_\omega^{PGL}/\mathcal{R}_0^{PGL}$  as a function of collision frequency; dashed line corresponds to  $\max \mathcal{R}_\omega^{PGL}/\mathcal{R}_0^{PGL} = Q^2 \propto \nu^{-2}$ .

$$\begin{aligned} \frac{\mathcal{R}_{(2n-1)\Omega_g}^{GL}}{\mathcal{R}_0^{GL}} &\simeq \left( \frac{1 + \zeta^{GL}}{1 + \eta^{GL}} \right) \\ &\times \frac{(\xi^{GL} Q)^2}{[(\xi^{GL} + Q)^2 + (2n - 1)^2 Q^4 (\pi \nu \tau^{GL} / 4)^2]} \end{aligned} \quad (44)$$

for GL-FETs, and

$$\frac{\mathcal{R}_{(2n-1)\Omega_g}^{PGL}}{\mathcal{R}_0^{PGL}} \simeq \frac{(\xi^{PGL} Q)^2}{[(\xi^{PGL} + Q)^2 + (2n - 1)^2 Q^4 (\pi \nu \tau^{PGL} / 4)^2]} \quad (45)$$

for PGL-FETs.

Using equation (44) and taking into account a long recharging time  $\tau^{GL}$  in the GL-FETs, for their responsivity in the range of several THz, we find

$$\begin{aligned} \frac{\mathcal{R}_{(2n-1)\Omega_g}^{GL}}{\mathcal{R}_0^{GL}} &\simeq \left( \frac{1 + \zeta^{GL}}{1 + \eta^{GL}} \right) \left( \frac{1}{\xi^{GL}} + \frac{1}{Q} \right)^{-2} \\ &\simeq \left( \frac{T}{2\epsilon_F} \right) \left( \frac{1}{\xi^{GL}} + \frac{1}{Q} \right)^{-2}. \end{aligned} \quad (46)$$

Thus, the height of the GL-FET detector responsivity maxima is determined by  $\min\{\xi, Q\}$ . Substituting to equation (46) the data used for the dependences shown in figures 3 and 4, we naturally obtain the same values of the peak height as in these plots.

Since the quantity  $\xi^{PGL}$  can be much larger than  $Q$ , for the low-index resonances equation (45) can be simplified:

$$\frac{\mathcal{R}_{(2n-1)\Omega_g}^{PGL}}{\overline{\mathcal{R}}_0^{PGL}} \simeq \frac{Q^2}{1 + (2n-1)^2 Q^4 \left( \frac{\pi \nu \tau^{PGL}}{4 \xi^{PGL}} \right)^2} \leq Q^2. \quad (47)$$

The dependence  $\mathcal{R}_{(2n-1)\Omega_g}^{PGL}/\overline{\mathcal{R}}_0^{PGL} \simeq Q^2$  at the realistic parameters is also confirmed by the proximity of the solid and dashed lines in the inset in figure 8.

Comparing equations (46) and (47), one can find that the responsivity peaks of the GL- and PGL-FET detectors (1) exhibit rather different height and sharpness and (2) change variously with increasing  $\varepsilon_F$ .

First, assuming that in reality  $\xi^{GL} \leq Q \ll \xi^{PGL}$ , disregarding for simplicity the "recharging" terms in the denominators of equations (44) and (45), taking into account equations(29), we find

$$\frac{\mathcal{R}_{(2n-1)\Omega_g}^{PGL}}{\mathcal{R}_{(2n-1)\Omega_g}^{GL}} \simeq \left( \frac{2\varepsilon_F}{T} \right)^2 \left( \frac{Q}{\xi^{GL}} \right)^2 \gg 1.$$

Second,  $\mathcal{R}_{(2n-1)\Omega_g}^{GL}$  decreases with increasing  $\varepsilon_F$ , while  $\mathcal{R}_{(2n-1)\Omega_g}^{PGL}$  increases. Indeed, according to Eq. (45) with Eqs. (11) and (34) at large values of  $\varepsilon_F$  and  $Q$ , one obtains  $\mathcal{R}_{(2n-1)\Omega_g}^{GL} \propto (T/\varepsilon_F)^{3/2}$ . Such a decreasing dependence on  $\varepsilon_F$  is associated with both a decrease in the tunneling nonlinearity parameter and in an increase in the p-n junction conductance. In contrast,  $\mathcal{R}_{(2n-1)\Omega_g}^{PGL} \propto Q \propto \Omega_g \propto \sqrt{\varepsilon_F}$ . Different behavior of the resonant peaks clearly seen comparing figures 5 and (6). A sensitivity of the responsivity maxima to the Fermi energy implies an opportunity to optimize the PGL-FET detector characteristics.

The peaks heights can markedly roll-off with increasing resonance index  $n$ , especially if  $\nu \tau^{PGL}$  is not too small. One needs to mention that  $\tau^{PGL}/\xi^{PGL} \propto 1/\sqrt{\varepsilon_F}$  and the second term in the denominator in equation (45) is approximately proportional to  $\varepsilon_F$ , whereas both these quantities are independent of  $\Delta_g$ . At very high Fermi energies, the dependence  $\mathcal{R}_{\omega}^{PGL}/\overline{\mathcal{R}}_0^{PGL}$  versus  $\varepsilon_F$  saturates and results in the limit

$$\frac{\mathcal{R}_{(2n-1)\Omega_g}^{PGL}}{\overline{\mathcal{R}}_0^V} \rightarrow \frac{1}{(2n-1)^2} \left( \frac{L_g}{2W_g} \right)^2.$$

### VIII. COMMENTS

In the above calculations we disregarded possible frequency dispersion of the p-n junction conductance, which actually contain the factor  $J_0(\omega t_{tr}/2)$ , where  $J_0(\tau)$  is the Bessel function and  $t_{tr} = l/v_W$  (in the GL-FETs) and  $t_{tr} = (l/v_W)\sqrt{\Delta_g/2T}$  (in the PGL-FETs) are the transit times across the depletion region. Such a transit-time

effect might be important at the frequencies commensurable with the inverse hole and electron time when the argument of the Bessel function  $J_0(\omega t_{tr}/2)$  is about of 2 or more). At  $\omega/2\pi = 1-3$  THz,  $l = 25$  nm, and  $\Delta_g = 200$  meV, one obtains  $\omega t_{tr} = 0.15 - 0.45$ . Hence, the assumption that  $\text{Re } g^{pn}$  is independent of the frequency used in our model is valid.

The incident THz radiation heats the 2DHS and 2DES due to the Drude and the interband absorption. The variation of the hole and electron effective temperature can lead to the dc source-drain current if the device is asymmetric. The latter can be achieved by applying of the bias voltage. This effect can be used for the so-called photothermoelectric detection [52, 54]. However, in the case of symmetric structure in the absence of the dc bias (considered above), the effective hole and electron temperatures in the pertinent sections of the channel are equal, and, hence, the photothermoelectric effect can be disregarded (in contrast to the voltage-biased p-n junctions [52]).

Apart from using the engineered band gap (as in the PGL-FETs under consideration), an effective suppression of the interband tunneling in the GL-FET p-n junction can also be achieved by using the saw-shaped gate electrodes instead of conventional bar-like gate electrodes [64–67]. In such a case, the dramatically reinforced hole and electron backscattering from the barrier in the p-n junction can lead to a substantial drop in the tunneling current and the p-n junction net conductance. The GL-FET detectors with the saw-shaped gate electrodes are beyond the scope of the present paper, need a separate consideration, and will be considered elsewhere.

### IX. CONCLUSIONS

In summary, we considered the proposed resonant THz detectors using FETs with the split gates, which electrically induce the lateral p-n junction in uniform and perforated GLs, GL-FET and PGL-FET detectors. The gated regions of the GL- and PGL-FETs play the role of plasmonic resonant cavities whereas the p-n junction depletion GL or PGL region enable the nonlinearity of the current-voltage characteristics. We calculated the responsivity the spatial distributions of the ac and rectified components of electric potential thermionic and tunneling currents caused by the incident THz radiation. Relating the rectified voltage component and the intensity of the THz radiation, we obtained formulas for the responsivity of the GL-FETs and PGL-FETs operating as the photovoltaic THz detectors. As shown, the spectral dependences of the responsivity can exhibit sharp resonant maxima associated with the excitation of plasmonic oscillations. The height of the responsivity peaks is determined by the hole and electron collision frequency in the channel section and by the relative conductance of the p-n junction. Due to a suppression of the tunneling current in the PGL-FETs (with the perforated p-

n-junction), their resonant response is much more pronounced and the responsivity is substantially higher than that in GL-FETs. The PGL-FET detectors operating at room temperature can be of interest for different THz communication systems.

### Acknowledgments

One of the authors (VR) thanks D. Svintsov for useful discussion and N. Ryabova for assistance. The work was supported by the Japan Society for Promotion of Science (Grant GA-SR-A #16H02336) and the Russian Scientific Foundation (Project #14-29-00277). The works at RPI was supported by the US Army Research Laboratory Cooperative Research Agreement.

### Appendix A. Differential conductance and nonlinearity parameter in GL-FETs

**Thermionic (activation) current.** Both the thermionic transfer of the carriers and their tunneling between the gate sections enable the nonlinear current voltage characteristics of the p-n junction. Using the Landauer-Buttiker formula [68] for 2D channels with the barrier transparency is given

$$\mathcal{T}(\varepsilon, V^{pn}) = \Theta(\varepsilon - 2\varepsilon_F + eV^{pn}), \quad (\text{A1})$$

where  $\Theta(\varepsilon)$  is the unity step function, and considering that the thermionic current is created by both the injected electron and hole components, for the thermionic current density one can obtain

$$j^{th} = J_s^{GL} \left[ \exp\left(\frac{eV^{pn}}{T}\right) - 1 \right], \quad (\text{A2})$$

Here  $V^{pn}$  is the signal voltage across the p-n junction, which includes the ac and rectified components,  $\delta V_\omega^{pn}$  and  $\overline{\Delta V}^{pn}$ , respectively, and the saturation current density  $J_s$  is given by

$$J_s^{GL} \simeq \frac{4eT^2}{\pi^2 \hbar^2 v_W} \exp\left(-\frac{\varepsilon_F}{T}\right) \quad (\text{A3})$$

Here  $T$  is the electron and hole temperature in the channel (in the energy units) and has been taken into account that the built-in voltage  $V^{bi} = 2\varepsilon_F/e$ .

As a result, at  $|V^{pn}| \ll T/e$ , from Eqs. (A1) and (A2) calculating  $g^{th} = (dj^{th}/dV)|_{V^{pn}=0}$  and  $\beta^{th} = \frac{1}{2}(d^2 j^{th}/dV^2)|_{V^{pn}=0}$ , we arrive at Eqs. (9) and (10).

**Tunneling current.** Considering that the net electric field in the p-n junction is equal to  $E^{bi} + E^{pn}$ , where  $E^{bi}$  and  $E^{pn}$  are the electric fields produced built-in voltage and the signal electric field component, the tunneling

current density in the GL-FETs, which is proportional to  $\sqrt{(E^{bi} + E^{pn})}|_{x=0}} V$  [1, 4, 8], can be presented as

$$j^{tunn} = g_0^{tunn} V^{pn} \sqrt{(E^{bi} + E^{pn})}|_{x=0}}. \quad (\text{A4})$$

Here

$$g_0^{tunn} = \left(\frac{e^2}{\pi^2 \hbar}\right) \sqrt{\frac{eE^{bi}}{\hbar v_W}}. \quad (\text{A5})$$

If the spacing between the gate edges  $2l \leq W_g$ , the electric potential in the region  $-l < x < l$  can be presented as  $\varphi \simeq [(V^{bi} + V)/2l]x$ , where  $V^{bi} = 2\varepsilon_F/2e$ . Hence,

$$E^{bi} + E \simeq \frac{V^{bi} + V^{pn}}{2l}. \quad (\text{A6})$$

When  $2l > W_g$ , neglecting the electron and hole space charges in the region  $-l < x < l$  and taking into account the smallness of the screening length  $r_S = \kappa \hbar^2 v_W^2 / 4e^2 T \ln[1 + \exp(\varepsilon_F/T)] \leq 10$  nm [21, 24, 69], the electric-field spatial distribution in this region is given by [8, 59]:

$$E^{bi} + E^{pn} = \frac{V^{bi} + V^{pn}}{\pi \sqrt{(l^2 - x^2)}}. \quad (\text{A7})$$

Therefore,  $(E^{bi} + E^{pn})|_{x=0} = (V^{bi} + V^{pn})/\pi l$ . The electron and hole space charge decrease the electric field in the p-n junction center [8, 21, 60], so that one can put  $(E^{bi} + E)|_{x=0} = (V^{bi} + V)/\pi l^* \leq (V^{bi} + V)/\pi l$ , where  $l^* \geq l$ . Considering this, we find that  $g^{tunn} = (dj_0^{tunn}/dV)|_{V^{pn}=0}$  and  $\beta^{tunn} = \frac{1}{2}(d^2 j_0^{tunn}/dV^2)|_{V^{pn}=0}$  are equal to

$$g^{tunn} \simeq g_0^{tunn} = \left(\frac{e^2}{\pi^2 \hbar}\right) \sqrt{\frac{2\varepsilon_F}{\pi \hbar v_W l^*}}, \quad (\text{A8})$$

$$\beta^{tunn} \simeq \frac{g_0^{tunn}}{2V^{bi}} = \left(\frac{e^2}{\pi^2 \hbar}\right) \frac{e}{2\sqrt{2}\pi \hbar v_W \varepsilon_F l^*}, \quad (\text{A9})$$

respectively.

From equations (A8) and (A9) for the parameters  $\eta^{GL} = g^{tunn}/g^{th}$  and  $\zeta^{GL} = \beta^{tunn}/\beta^{th}$  defined by equation (8) in the main text, considering equations (9) and (10), we obtain equations (11) and (12).

### Appendix B. Differential conductance and nonlinearity parameter in PGL-FETs

**Thermionic injection: sharp barrier.** If the barrier in the depletion layer associated with the nanoconstrictions is sufficiently sharp, its shape can be approximated by a trapezoid. In this case, equation (A1) should be replaced by

$$j^{th} = J_s^{PGL} \left[ \exp\left(\frac{eV^{pn} - \Delta_g/2}{T}\right) - 1 \right], \quad (B1)$$

where  $J_s^{PGL} = 4eNb/\pi\hbar$ ,  $N$  is the density of the nanoconstrictions, and  $\Delta_g$  is the energy gap in the nanoconstrictions. Calculating  $(dj^{th}/dV)|_{V^{pn}=0}$  and  $(d^2j^{th}/dV^2)|_{V^{pn}=0}$ , from equation (B1) we get equations (13) and (14) which accounts for the 1D transport in the nanoconstrictions and the exponential factor  $\exp(-\Delta_g/2T)$  inherent to the activation processes.

**Tunneling current: sharp barrier.** Assuming that the hole and electron dispersion lows in the constrictions (GNRs) are  $\varepsilon^\pm = \mp\sqrt{p^2v_W^2 + (\Delta_g/2)^2} \simeq \mp[(\Delta_g/2) + p^2/2m]$  (where  $m = \Delta_g/2v_W^2$ ) and using the Landauer-Buttiker formula [68] for 1D channels with the tunneling transparency  $\mathcal{T}(\varepsilon, V^{pn})$  of the trapezoid barrier in the PGL-FET perforated region derived in the WKB approximation, we find

$$j^{tunn} \propto 2eN \times \exp\left\{-\frac{4\sqrt{m\varepsilon_F}l}{3\hbar} \left[ \left(\frac{\Delta_g}{2\varepsilon_F} + 1\right)^{3/2} - \left(\frac{\Delta_g}{2\varepsilon_F} - 1\right)^{3/2} \right] \right\} \quad (B2)$$

(for  $\Delta_g/2\varepsilon_F = Z \geq 1$ ). Consequently,

$$g^{tunn} \simeq \frac{ej^{tunn}}{\varepsilon_F}, \quad \beta^{tunn} \simeq \frac{e^2j^{tunn}}{2\varepsilon_F^2}, \quad (B3)$$

Comparing the thermionic and tunneling contributions, one can find that in the case of the trapezoid tunneling barrier ( $Z = \Delta_g/2\varepsilon_F \geq 1$ ), the former contribution substantially exceed the tunneling contribution (so that  $\eta^{PGL}$  and  $\zeta^{PGL}$  are much smaller than unity) when

$$l \gg l^{tunn} \frac{1+Z}{\sqrt{Z}[(Z+1)^{3/2} - (Z-1)^{3/2}]} \quad (B4)$$

$l^{tunn} = 3\hbar v_W/4T$ . At room temperature  $l^{tunn} \simeq 18$  nm. If  $Z = 2 - 3$ , the tunneling can be disregarded in the PGL-FETs with  $l \gg 7 - 8$  nm.

Since  $g^{tunn}$  and  $\beta^{tunn}$  in the PGL-FETs are proportional to the tunneling current, which is weak, we can disregard their contribution to  $g^{pn}$  and  $\beta^{pn}$  [see equation (8)], keeping in mind that  $\zeta^{PGL}$  and  $\eta^{PGL}$  are exponentially small.

**Thermionic and tunneling currents: parabolic barrier potential model.** In the PGL-FET device structure, the shapes of the barriers for holes and electrons can also be approximated by parabolic functions [70, 71], so that in the barrier ( $|x| < l$ ) one has  $\Delta_g(x) = \Delta_g(1 - x^2/l^2)$ , where in such a case  $\Delta_g$  is the energy gap in the constriction center. Assuming the latter and using the Landauer-Buttiker formula [72], (adopted for the case of 1D transport in the nanoconstrictions under consideration) in which the barrier transparency is given by the Kemble formula [68]

$$\mathcal{T}(\varepsilon, V^{pn}) = \left[ 1 + \exp\left(\frac{\Delta_g^*/2 - \varepsilon}{\Theta}\right) \right]^{-1}, \quad (B5)$$

where

$$\Delta_g^* = \Delta_g \left( 1 - \frac{eV^{pn}}{2\Delta_g} \right)^2 \simeq \Delta_g - eV^{pn} \quad (B6)$$

is the maximum barrier height with respect to the bottom of the conduction band in the n-section. The "tunneling" temperature is given by

$$\Theta = \frac{\hbar}{2\pi} \sqrt{\frac{\Delta_g}{l^2 m}} \sim \frac{\hbar}{2\pi} \frac{v_W}{l} \quad (B7)$$

characterizes the contributions of the transition below the barrier top, where  $m$  is an effective mass in the barrier. Setting  $l = 25 - 100$  nm, we obtain  $\Theta \sim 11.5 - 46.0$  K. It implies that for room temperature one can assume  $\Theta \ll T$ . Taking the latter into account, for the PGL-FETs with not too wide constrictions (in which  $\Delta_g \gg 2\Theta$ ), we again arrive at equations (13) and (14) with  $\zeta^{PGL} < \eta^{PGL} \ll 1$ . However, if  $l = 7 - 8$  nm,  $\Theta \sim 390 - 450$  K. In this case, the tunneling processes can be marked. This is in line with the results of the above trapezoidal barrier model.

---

[1] Cheianov V V and Fal'ko V I 2006 Phys. Rev. B **74** 041403(R)  
[2] Ossipov A, Titov M, and Beenakker C W.J 2007 Phys. Rev. B **75** 241401'  
[3] Zhang L M and Fogler M M 2008 Phys. Rev. Lett. **100** 116804  
[4] Low T, Hong S, Appenzeller J, Datta S and Lundstrom M S 2009 IEEE Trans. Electron Devices, **56** 1292  
[5] Ryzhii M, Ryzhii V, Otsuji T, Mitin V and Shur M S

20110 Phys. Rev. B **82** 075419  
[6] Liu L, Safavi-Naeini S and Ban D 2014 Electronics Lett, **50** 1724  
[7] Ryzhii V, Ryzhii M, Mitin V and Shur M S 2009 Appl. Phys. Express **2** 034503  
[8] Semenenko V L, Leiman V G, Arsenin A V, Mitin V, Ryzhii M, Otsuji T and Ryzhii V 2013 J. Appl. Phys. **113** 024503  
[9] Xia F, Mueller T, Lin Y-M, Valdes-Garcia A and Avoris

- P 2009 Nature Nanotech, **4** 839
- [10] Ryzhii V, Ryzhii M, Mitin V and Otsuji T 2010 J. Appl. Phys. **107** 054512
  - [11] Mueller T, Xia F and Avouris P 2010 Nature Photon. **4** 297
  - [12] Ryzhii M, Otsuji T, Mitin V and Ryzhii V 2011 Jpn. J. Appl. Phys. **50** 070117
  - [13] Gabor N M, Song J C W, Ma Q, Nair N L, Taychatanapat T, Watanabe K, Taniguchi T, Levitov L S and Jarillo-Herrero P 2011 Science, **334** 648
  - [14] Ryzhii V, Ryabova N, Ryzhii M, Baryshnikov N V, Karasik V E, Mitin V. and Otsuji T 2012 Optoelectronics Rev. **20** 15.
  - [15] Sun D, Aivazian G, Jones A M, Ross J, Yao W, D. Cobden D and Xu X 2012 Nature Nanotechnology **7** 114
  - [16] Freitag M, Low T and Avouris P 2013 Nano Lett. **13** 1644
  - [17] Koppens F H L, Mueller T, Avouris Ph, Ferrari A C, Vitiello M S and Polini M 2014 Nature Nanotech. **9** 780
  - [18] Baugher B W H, Churchill H O H, Yang Y and Jarillo-Herrero P 2014 Nature Nanotech. **9** 262
  - [19] Ryzhii M and Ryzhii V 2007 Jpn. J. Appl. Phys. **46** L151
  - [20] Ryzhii V, Ryzhii M, Mitin and Otsuji T 2011 J. Appl. Phys. **110** 094503
  - [21] Ryzhii V, Semenikhin I, Ryzhii M, Svintsov D, Vyurkov V, Satou A and Otsuji T 2013 J. Appl. Phys. **113** 244505
  - [22] Tamamushi G, Watanabe T, Dubinov A A, Wako H, Satou A, Suemitsu T, Ryzhii M, Ryzhii V and Otsuji T 2016 CLEO: Conf. on Lasers and Electro-Optics, SM3L.7, San Jose, CA, USA, 6 June 2016. (to be presented)
  - [23] Lu Ye, Goldsmith B, Strachan D R, Lim J, Luo Z and Johnson A T C 2010 Small **6** 2748
  - [24] H. Mohamadpour H 2012 Physica E **46** 270
  - [25] Naderi A 2015 drain regions, J. Comput. Electron. **16** 1
  - [26] Zhang L M and Fogler M M 2008 Phys. Rev. Lett. **100** 116804
  - [27] Fogler M M, Novikov D S, Glazman L I and Shklovskii B I 2008 Phys. Rev. B **77** 075420
  - [28] Liu J, Safavi-Naeini S and Ban D 2014
  - [29] Zhoua Y *et al.* 2016 Phys. Lett. A **380**, 1049
  - [30] Miyazaki H, Lee M V, Li S-L, Uira H, Kanda A and Tsukagoshi K 2012 "J. Phys. Soc. Jpn. **81** 014708
  - [31] Aparecido-Ferreira A, Miyazaki H, Li S-L, Komatsu K, Nakaharai S and Tsukagoshi K 2012 Nanoscale, **4** 7842
  - [32] Nandkishore R and Levitov L 2011 Proc. Nat. Acad. Sci. USA **108** 14021
  - [33] Hung Nguyen V, Saint-Martin J, Querlioz D, Mazzamuto F, Bournel, Niquet Y-M and Dollfus P 2013 J. Comput. Electron. **12** 85
  - [34] Dyakonov M I and Shur M S 1996 IEEE Trans. Electron Devices, **43** 1640
  - [35] Knap W *et al.* 2009 J. Infrared Millimeter Terahertz Waves, **30** 1319
  - [36] Nadar S *et al.* 2010 J. Appl. Phys. **108** 054508
  - [37] Popov V V, Fateev D V, Otsuji T, Meziani Y M, Coquillat D and Knap W 2011 Appl. Phys. Lett. **99** 243504
  - [38] Kurita Y *et al.* 2014 Appl. Phys. Lett. **104** 251114
  - [39] S.A. Boubanga Tombet S A *et al.* 2014 Appl. Phys. Lett. **104** 262104
  - [40] Castro Neto A H, Guinea F, Peres N M R, Novoselov K S and Geim A K 2009 Rev. Mod. Phys. **81** 109
  - [41] Das Sarma S, Adam S, Hwang E H and Rossi E 2011 Rev. Mod. Phys. **83** 407
  - [42] Bolotin K I, Sikes K J, Z. Jiang Z, M. Klima M, Fudenberg G, Hone J, Kim P and Stormer H L 2008 Solid State Comm. **146** 351
  - [43] Tamamushi G, Sugawara K, Binti Hussin M, Suemitsu T, Suto R, Fukidome H, Suemitsu M and Otsuji T 2015 Comp. Semicond. Week 2015, Santa Barbara, CA. USA, June 28-July 2
  - [44] Vicarelli L, Vitiello M S, Coquillat D, Lombardo A, Ferrari A C, Knap W, Polini M, Pellegrini V and Tredicucci A 2012 "Nature Mat. **11** 865
  - [45] Spirito D *et al.* 2014 Appl. Phys. Lett. **104** 061111
  - [46] Wang H, Nezich D, Kong J and Palacios T 2009 IEEE Electron Device Lett. **30** 547
  - [47] Chen H-Y and Appenzeller J 2012 Nano Lett. **12** 2067
  - [48] Ryzhii V, Satou A and Otsuji T 2007 J. Appl. Phys. **101** 024509
  - [49] Ryzhii V, Satou A, Otsuji T, Ryzhii M, Mitin V and Shur M S 2013 J. Phys. D: Appl. Phys. **46** 315107.
  - [50] Ryzhii V, Otsuji T, Ryzhii M, Mitin V and Shur M S 2015 J. Appl. Phys. **118** 204501
  - [51] Wang X, Quyang Y, Li X, Wang H, Guo J, and Dai H 2008 Phys. Rev. Lett. **100** 206803
  - [52] Ryzhii V, Otsuji T, Ryzhii M, Ryabova N, Yurchenko S O, Mitin V and Shur M S 2013 J. Phys. D **46**, 065102
  - [53] Watanabe T, Fukushima T, Boubanga Tombet S, Satou A, Otsuji T, Chan S, Ryzhii V and Popov V 2011 Ext. Abstr. of Int. TeraNano & GDRI Workshop, 28P-04, Osaka, Nov. 24-29
  - [54] Cai X, A. Sushkov A B, Jadidi M M, Nyakiti L O, Myers-Ward R L, Gaskill D K, Murphy T E, Fuhrer M S and Drew H D 2015 Nano Lett. **15** 4295
  - [55] Boubanga-Tombet S, Teppe F, Coquillat D, Nadar S, Dyakonova N, Videlier H and Knap W 2008 Appl. Phys. Lett. **92**, 212101
  - [56] Simin G S, Islam M, Gaevski M, Gashka R and Shur M S 2014 IEEE Electron Device Lett. **35** 449
  - [57] Shur M S 1990 *Physics of Semiconductor Devices* (Englewood Cliffs, NJ: Prentice Hall, New Jersey)
  - [58] S. Luryi S 1988 Appl. Phys. Lett. **52** 501
  - [59] Fang T, Konar A, Xing H L and Jena D 2007 Appl. Phys. Lett. **91** 092109
  - [60] Chklovskii D B, Shklovskii B I and Glazman L I 1992 Phys. Rev. B **46** 4026
  - [61] Frensley W R 1980 IEEE Electron Device Lett. **1** 137
  - [62] Svintsov D, Vyurkov V, Yurchenko S, Otsuji T and Ryzhii V 2013 J. Appl. Phys. **111** 083715
  - [63] Narozhny B N, Gornyi I V, Titov M, Schutt M and Mirlin A D 2015 Phys. Rev. B **91** 035414
  - [64] Jiang M S, Kim H, Son Y-W, Atwater H A and Goddard III W A 2013 Proc. Nat. Acad. Sci USA **110** 8786.
  - [65] Sajjad R N, Sutar S, Lee J U and Ghosh A W 2012 Phys. Rev. B **86** 155412
  - [66] Sutar S, Comfort E S, Liu J, Taniguchi T, Watanabe K and Lee J U 2012 Nano Lett. **12** 4460
  - [67] Wilmar Q, Berrada S, Torrin D, Hung Nguyen V, Fve G, Berroir J-M, Dollfus P and Plaais B 2014 2D Materials **1** 01106
  - [68] Buttiker M, Imry Y, Landauer R and Pinhas S 1985 Phys. Rev. B **31** 6207
  - [69] Das Sarma S and Hwang W H 2007 Phys. Rev. B **75** 205418
  - [70] Matveev K A and Glazman L I 1996 Phys. Rev. B **54** 10339
  - [71] Hansen K and Brandbyge M 2004 J. Appl. Phys. **95** 3582
  - [72] Kemble E C 1935 Phys. Rev. **48** 549